Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Preliminary crystallographic analysis of the NAC domain of ANAC, a member of the plant-specific NAC transcription factor family

Addie Nina Olsen,[a]† Heidi Asschenfeldt Ernst,[b]† Leila Lo Leggio,[b]* Eva Johansson,[b,c] Sine Larsen[b,c] and Karen Skriver[a]

[a]Institute of Molecular Biology, University of Copenhagen, Øster Farimagsgade 2A, DK-1353 Copenhagen K, Denmark, [b]Centre for Crystallographic Studies, Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark, and [c]European Synchrotron Radiation Facility (ESRF), 6 Rue Jules Horowitz, 38000 Grenoble, France

† These authors contributed equally to the work.

Correspondence e-mail: leila@ccs.ki.ku.dk

Received 7 August 2003
Accepted 3 October 2003

The NAC domain (residues 1–168) of ANAC, encoded by the abscisic acid-responsive NAC gene from *Arabidopsis thaliana*, was recombinantly produced in *Escherichia coli* and crystallized in hanging drops. Three morphologically different crystal forms were obtained within a relatively narrow range of conditions: 10–15% PEG 4000 and 0.1 *M* imidazole/malic acid buffer pH 7.0 in the reservoir, 3.2–7.7 mg ml^{-1} protein stock and a 1:1 ratio of reservoir to protein solution in the hanging drop. One of the crystal forms, designated crystal form III, was found to be suitable for further X-ray analysis. Form III crystals belong to space group $P2_12_12_1$, with unit-cell parameters a = 62.0, b = 75.2, c = 80.8 Å at 100 K. The unit-cell volume is consistent with two molecules in the asymmetric unit and a peak in the native Patterson map suggests the presence of a non-crystallographic twofold axis parallel to a crystallographic axis. Size-exclusion chromatography of the NAC domain showed that the dimeric state is also the preferred state in solution and probably represents the biologically active form. Data sets were collected from four potential heavy-atom derivatives of the form III crystals. The derivatized crystals are reasonably isomorphous with the non-derivatized crystals and the four data sets are being evaluated for use in structure determination by multiple isomorphous replacement.

1. Introduction

The development and survival of living organisms depend on proper regulation of gene expression and consequently on a myriad of transcription factors. The regulatory action of transcription factors involves interaction with specific DNA sequences. Thus, determination of the structure of the DNA-binding domains of transcription factors is essential in order to understand their mechanism of action.

Information on the structure of plant transcription factors is often obtained by homology modelling based on the structures of animal transcription factors. However, the existence of plant-specific families of transcription factors necessitates a more direct approach. NAC proteins are a recently identified family of plant-specific transcriptional regulators (Souer *et al.*, 1996; Aida *et al.*, 1997). NAC protein-encoding genes constitute large gene families in a wide range of plant species. For example, the genome of the model organism *Arabidopsis thaliana* contains at least 90 NAC genes (Duval *et al.*, 2002). Members of the NAC protein family have crucial and diverse functions in plant biology. For example, NAC proteins are involved in the development of the shoot apical meristem and floral organs and in the formation of lateral roots (Souer *et al.*, 1996; Aida *et al.*, 1997; Xie *et al.*, 2000). Furthermore, NAC proteins have been implicated in the response to pathogens (Xie *et al.*, 1999; Ren *et al.*, 2000; Collinge & Boller, 2001). The protein family is characterized by a highly conserved N-terminal domain, known as the NAC domain, whereas the C-terminal regions are diverse (Kikuchi *et al.*, 2000). Three *Arabidopsis* NAC proteins, TIP, NAC1 and NAM, have been shown to function as transcription activators. The DNA-binding ability of NAC1 and NAM has been assigned to the NAC domain; the C-terminal regions of both proteins constitute transcriptional activation domains (Ren *et al.*, 2000; Xie *et al.*, 2000; Duval *et al.*, 2002). NAC proteins have also been shown to interact with different proteins, including viral proteins (Ren *et al.*, 2000; Xie *et al.*, 1999) and RING finger proteins (Xie *et al.*, 2002; Greve *et al.*, 2003). ANAC is a recently identified *Arabidopsis* NAC protein that interacts with a RING protein (Greve *et al.*, 2003). The ANAC gene is regulated by the plant stress hormone abscisic acid and therefore represents an addition to the increasing subgroup of hormone-inducible NAC genes (Xie *et al.*, 2002; Hoth *et al.*, 2002).

The three-dimensional structure of an NAC protein has yet to be determined. A far-UV circular-dichroism study of the recombinant NAC domain from ANAC indicated its secondary structure to be dominated by

β-strands (Greve *et al.*, 2003). Based on secondary-structure predictions, others have suggested that the NAC subdomains responsible for interaction with DNA form a helix–turn–helix structure (Duval *et al.*, 2002). In order to improve our understanding of NAC-protein function, we have initiated the structure determination of the recombinant NAC domain from ANAC. Here, we report the crystallization of this domain, native data collection and data collection from four potential heavy-atom derivatives for use in phasing.

2. Cloning, expression and purification

The DNA fragment encoding the NAC domain of ANAC (accession No. NP_175697, residues 1–168) was cloned into expression vector pET-15b (Novagen) as described previously (Greve *et al.*, 2003). The recombinant protein produced from this construct was soluble, in contrast to several other ANAC fragments produced initially, suggesting that it defines the NAC domain. *Escherichia coli* strain BL21(DE3) was transformed with the construct and used for expression of the His-tagged recombinant protein. A starter culture of 10 ml LB medium containing 100 µg ml^{-1} ampicillin was inoculated with the transformants and grown overnight at 310 K to stationary phase. 3 ml starter culture was transferred to 300 ml LB/ampicillin medium. The culture was grown at 310 K to a cell density (OD_{600}) of 0.6 before isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 m*M*. After 3 h of incubation at 310 K, the cells were harvested by centrifugation and stored at 253 K.

The cell pellet was resuspended in 24 ml 50 m*M* sodium phosphate pH 7.0, 0.3 *M* NaCl and lysozyme (Fluka BioChemika) was added to a concentration of 0.75 mg ml^{-1}. Following a 20 min incubation at room temperature, the cells were disrupted by sonication. The lysate was cleared by centrifugation at 10 000*g* for 20 min at 277 K and bound to 1.5 ml prewashed BD Talon metal-affinity resin (Clontech). The resin was washed twice in washing buffer (50 m*M* sodium phosphate pH 7.0, 0.5 *M* NaCl, 10% glycerol, 10 m*M* imidazole) followed by two washes with thrombin cleavage buffer (20 m*M* Tris–HCl pH 7.4, 150 m*M* NaCl). The recombinant NAC domain (with three additional N-terminal amino acids from the His-tag spacer region) was separated from the His tag by incubating the resin with 100 U thrombin protease (Amersham Biosciences) for 2 h at room temperature. The NAC protein released by the cleavage reaction was washed out of the column with 20 m*M* Tris–HCl pH 7.4, 0.5 *M* NaCl and passed through a 1 ml HiTrap Benzamidine FF column in order to remove the protease. Subsequently, a concentration and buffer change to 20 m*M* Tris–HCl pH 7.5 were performed using Amicon Ultra-15 centrifugal filter units (Millipore). Typically, 3–4 mg of pure protein were obtained from a 300 ml preparation.

3. Oligomerization state of the NAC domain

Size-exclusion chromatography was performed on an ÄKTA Purifier 10 system using a Superdex 75 HR 10/30 column (Amersham Biosciences). The NAC domain was purified as described above, except that the buffer was exchanged with 20 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl following protease removal. The protein preparation was filtered through an Ultrafree-MC 0.45 µm centrifugal filter unit (Millipore) and a 150 µl sample at a concentration of approximately 1.6 mg ml^{-1} was applied to the column. The column was eluted with 20 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl at a flow rate of 0.5 ml min^{-1}. The chromatographic separation was calibrated using as standards the following proteins: bovine serum albumin (66 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), myoglobin (17.6 kDa) and cytochrome *c* (12.4 kDa). As seen in Fig. 1, most of the NAC domain elutes as a peak centred at an elution volume of 10.54 ml. The peak is thus positioned between the standard protein peaks corresponding to ovalbumin (43 kDa) and carbonic anhydrase (29 kDa). The theoretical molecular weight of an ANAC NAC domain is 19.75 kDa and the major peak in the elution profile is therefore most consistent with a dimeric form of the protein. A small proportion of the NAC domain elutes close to the myoglobin standard (17.6 kDa) and is likely to correspond to the monomeric form. Thus, we conclude that NAC is preferentially present in solution as a dimer, with a small proportion present as the monomer. Similar results were obtained with a concentration of 5 mg ml^{-1}. The ability of the NAC domain to dimerize has previously been observed in a yeast two-hybrid assay (Xie *et al.*, 2000).

4. Crystallization

Crystals were grown at room temperature by the hanging-drop vapour-diffusion method in Linbro plates using protein stocks of 3.2–7.7 mg ml^{-1} in drops consisting of 2 µl protein and 2 µl reservoir solution equilibrated over 500 µl reservoirs. Initial solubility screening (Stura *et al.*, 1991) resulted in crystals under several conditions and a variety of morphologically different crystal forms were subsequently identified from Hampton Research Crystal Screen I (Jancarik & Kim, 1991) and preliminary optimization experiments. Generally, polyethylene glycols (PEGs) of high molecular

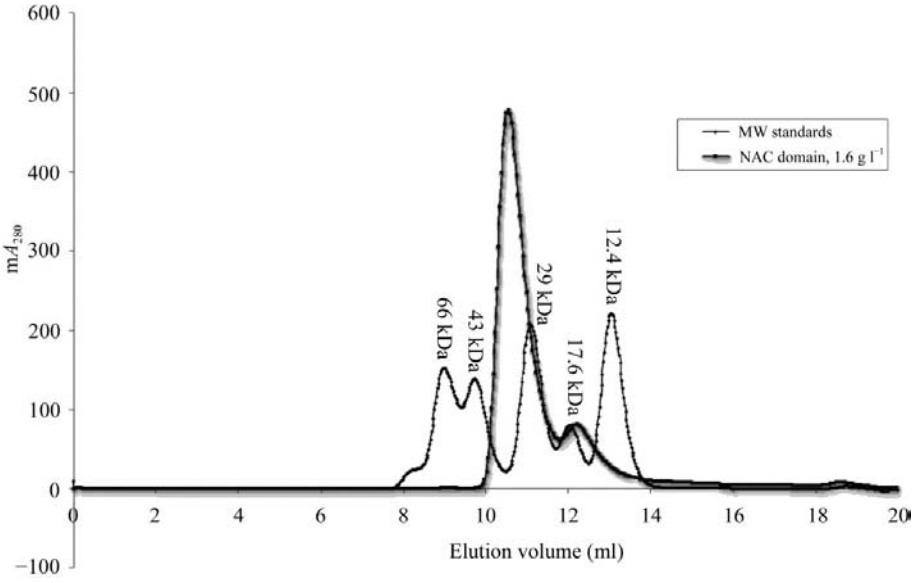

Figure 1
Size-exclusion chromatography of the ANAC NAC domain. The major NAC-domain peak elutes just before the cabonic anhydrase standard (29 kDa), while a minor peak in the sample almost coincides with the myoglobin peak (17.6 ka). The molecular weights of the standard proteins are shown in the figure.

Table 1
Data-collection statistics for form III crystals of ANAC(1–168).

Values in parentheses refer to the highest resolution shell.

	Native	$KAu(CN)_2$	TMLA	EMTS	K_2PtCl_4
Beamline	I711, MAXLAB	ID29, ESRF	ID29, ESRF	ID29, ESRF	ID29, ESRF
Resolution range (Å)	30–2.50 (2.64–2.50)	30–1.90 (2.00–1.90)	30–3.00 (3.16–3.00)	30–3.00 (3.16–3.00)	30–3.50 (3.69–3.50)
Completeness (%)	95.6 (89.8)	97.9 (95.1)	99.9 (99.9)	99.9 (99.9)	99.3 (99.3)
Unique reflections	11476	26725	8008	7854	4797
Multiplicity	4.0 (4.0)	4.1 (2.8)	9.2 (9.1)	3.8 (3.9)	6.4 (6.7)
$\langle I/\sigma(I)\rangle$	5.6 (2.5)	6.1 (2.3)	8.7 (4.7)	6.8 (2.4)	5.1 (2.4)
$R_{\rm merge}$†	0.094 (0.300)	0.061 (0.307)	0.072 (0.154)	0.097 (0.296)	0.121 (0.266)
Scaling R factor		0.161 (10–3 Å)	0.258 (10–3 Å)	0.320 (10–3 Å)	0.280 (10–3.5 Å)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I(hkl)_i - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I(hkl)_i$, where $I(hkl)_i$ is the ith measurement and $\langle I(hkl)\rangle$ is the average intensity of symmetry-equivalent reflections.

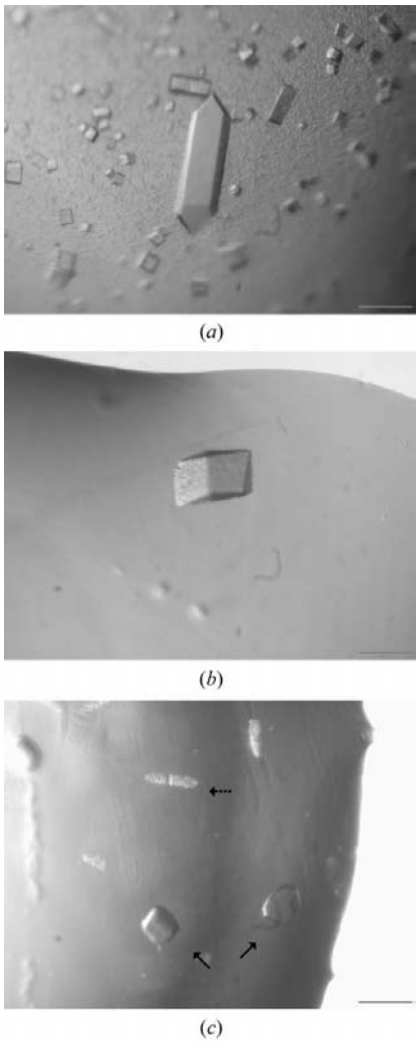

Figure 2
Crystals of ANAC (1–168). (*a*) Large form I crystal and many small crystals with different morphology, (*b*) form II crystal, (*c*) form III crystals (indicated with arrows) and 'aged' form I crystals (dashed arrow). The scale bar corresponds to 100 µm. The crystals shown and those used for diffraction analysis were grown using the following conditions: form I, 4.3 mg ml^{-1} protein, 15% PEG 4000, 0.1 *M* imidazole/malic acid buffer pH 7.0; form II, 3.2 mg ml^{-1} protein, 10% PEG 4000, 0.1 *M* imidazole/malic acid buffer pH 7.0; form III, 6.4 mg ml^{-1}, 15% PEG 4000, 0.1 *M* imidazole/malic acid buffer pH 7.0 and 5% glycerol.

weight (≥4000 Da) were successful precipitants over a broad pH range (pH 5.5–8.5).

During optimization trials, it became clear that morphologically different crystal forms grow within a relatively narrow range of conditions (10–15% PEG 4000 in 0.1 *M* imidazole/malic acid buffer pH 7.0) and often in the same drop. The protein concentration and the age of the stock seem to be critical parameters determining the form(s) obtained. Reasonably sized single crystals were grown of three different forms (I, II and III; Fig. 2), which were further characterized. The hexagonal rod-shaped form I crystals (Fig. 2*a*) formed very quickly, typically in 1 h to 1 d, and reached a maximum size of 0.30 × 0.08 × 0.08 mm, whereas the form II crystals and the form III polygonal plates/cubes (Figs. 2*b* and 2*c*) grew in 2–5 d to maximum dimensions of 0.12 × 0.10 × 0.10 mm and 0.09 × 0.08 × 0.05 mm, respectively. Signs of ageing, including cracking and/or loss of transparency, were often observed for the form I and II crystals, an example of which is shown in Fig. 2(*c*).

Interestingly, replacement of the imidazole/malic acid buffer in the crystallization mixture with *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES) failed to produce crystals, whereas the combination of HEPES buffer and 20 m*M* imidazole used as an additive gave crystals, but of a different crystal habit. This suggests that imidazole favours the crystallization of ANAC (1–168) through specific interactions.

Izit (Hampton Research) is a commercial dye commonly used to differentiate protein crystals from salt crystals, based on the principle that the dye will absorb into the solvent channels present in protein crystals. In this study, no colouring of the crystals could be observed when using Izit to test whether the first crystals obtained were of protein nature. However, it could not be excluded that the lack of staining might be caused by traces of low-molecular-weight contaminants, perhaps glycerol, remaining in the protein preparation. It has previously been shown that small organic molecules can prevent dye absorption by protein crystals (Eckert *et al.*, 2003). Transfer of a crystal to a new drop containing only reservoir solution and Izit resulted in rapid dye absorption by the crystal and emphasizes the fact that even trace amounts of low-molecular-weight contaminants might interfere with Izit staining of protein crystals.

5. Data collection

5.1. Preliminary X-ray characterization of the three crystal forms

Initial tests of the large form I crystals performed in-house using Cu $K\alpha$ radiation (λ = 1.5418 Å) showed limited diffraction power and potential problems with cryocooling. Data collected at 100 K using reservoir solution with 25% glycerol as cryoprotectant showed high mosaicity and a resolution limit of 7 Å, whereas data collected at 293 K from a similar crystal mounted in a capillary extended to approximately 4.5 Å. Indexing of the room-temperature test images with *DENZO* (Otwinowski & Minor, 1997) was consistent with a primitive hexagonal crystal system with unit-cell parameters $a = b = 69.8$, $c = 184.7$ Å, $\gamma = 120^\circ$. Further tests of the different crystal forms and cryocooling conditions were carried out using beamlines X11 and X13 at EMBL, Hamburg, Germany. The form I crystals did not diffract beyond 7 Å at cryogenic temperatures and the resolution limit for form II was around 4 Å, with poor diffraction quality in terms of spot shape. However, despite their small size, the form III crystals showed diffraction beyond 2.5 Å and were of suitable quality for further X-ray analysis. Of several cryocooling conditions tested, the best results were obtained by direct flash-freezing of form III crystals grown in the presence of 5% glycerol.

5.2. Native data

A data set extending to 2.5 Å was collected at 100 K from a form III crystal grown in the presence of 5% glycerol and flash-frozen in liquid nitrogen at I711, MAXLAB, Lund, Sweden. This beamline operated at a wavelength of 1.0350 Å and was equipped with a MAR Research 165 mm CCD detector. Based on systematic absences, the space group of the primitive orthorhombic cell could unambiguously be assigned as $P2_12_12_1$, with unit-cell parameters a = 62.0, b = 75.2, c = 80.8 Å. The

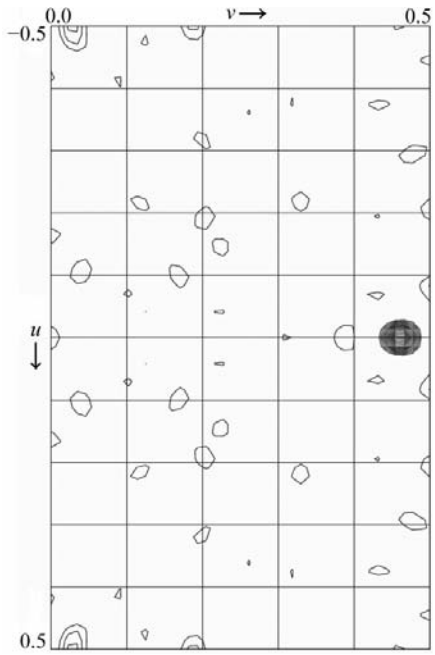

Figure 3
The native Patterson map reveals a peak (at $u = 0$, $v = 0.46$, $w = 0.5$) consistent with a non-crystallographic twofold axis parallel to a crystallographic axis. The $w = 0.5$ section of the map contoured in 1.5σ levels is shown.

unit-cell volume is consistent with two molecules in the asymmetric unit and a corresponding solvent content of 30% (V_M = 2.4 Å^3 Da^{-1}; Matthews, 1968). *MOSFLM* (v.6.2.2), *TRUNCATE* and *SCALA* from the *CCP*4 program suite v.4.2.2 (Collaborative Computational Project, Number 4, 1994) were used for processing, reduction and scaling of the diffraction data. Data-collection statistics are shown in Table 1. The native Patterson map calculated with the *CCP*4 program *FFT* (Fig. 3) revealed a strong peak at $u = 0$, $v = 0.46$, $w = 0.5$, which is consistent with a non-crystallographic twofold axis parallel with c through $x = 0.25$ and $y = 0.27$. The y coordinate being close to 0.25 makes the pseudosymmetry consistent with another twofold axis parallel with b through $x = 0.0$ and $z = 0.0$ with a small (0.04) translational component along y. Which of the two twofold axes relates the subunits of the functional dimer must be identified by structure determination.

5.3. Heavy-atom derivatives

Data sets from four putative heavy-atom derivatives were collected at 100 K on beamline ID29, ESRF, Grenoble, France to at least 3.5 Å resolution. An ADSC Q210 CCD detector was used. Crystals originating from the same drop as the native crystal were soaked for the times indicated in drops consisting of 1 µl heavy-atom solution and 4 µl reservoir solution, with final heavy-atom concentrations of 10 m*M* $KAu(CN)_2$ (9 h), 5 m*M* trimethyl lead acetate (TMLA, 6 h), 15 m*M* ethyl mercurithiosalicylate (EMTS, 9 h) and 2 m*M* K_2PtCl_4 (20 h), respectively, and flash-frozen directly without the addition of cryoprotectant. A wavelength of 0.9999 Å was selected in order to obtain anomalous contributions from all four heavy elements. The crystal soaked in K_2PtCl_4 for 20 h clearly deteriorated, even after a very short time in the heavy-atom-containing solution. Nevertheless, a full data set to 3.5 Å could be collected. On the contrary, the data obtained from the Au derivative extend to 1.9 Å, which is significantly better than the native data available. Processing of the derivative data sets with *MOSFLM* (v.6.2.2), *TRUNCATE* and *SCALA* and subsequent scaling to the native data using the *CCP*4 program *SCALEIT* showed that the soaked crystals were all isomorphous to the native crystal, as reflected in the average unit-cell parameters $a = 62.1 \pm 0.4$, $b = 75.1 \pm 0.3$, $c = 80.1 \pm 1.1$ Å and relatively low scaling R factors (Table 1).

6. Conclusions

Diffraction data have been obtained from a native crystal and four isomorphous potential heavy-atom derivatives of the NAC domain of ANAC. Preliminary runs of *SOLVE* (Terwilliger, 2002) suggest that all four heavy atoms are bound to the protein. Structure determination by multiple isomorphous replacement utilizing anomalous contributions is currently in progress.

We thank Flemming Hansen for help with data collection and MAXLAB (Lund, Sweden), the ESRF (Grenoble, France) and EMBL (Hamburg, Germany) for the provision of synchrotron beamtime and assistance. We are grateful to Dansync and the European Union under the 'Access to Research Infrastructure' program for contribution to travel expenses. This research was funded by the Danish National Research Foundation, a grant from the Danish Research Council to KS and a PhD stipend from University of Copenhagen to ANO.

References

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H. & Tasaka, M. (1997). *Plant Cell*, **9**, 841–857.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Collinge, M. & Boller, T. (2001). *Plant Mol. Biol.* **46**, 521–529.
Duval, M., Hsieh, T.-F., Kim, S. Y. & Thomas, T. L. (2002). *Plant Mol. Biol.* **50**, 237–248.
Eckert, K., Ernst, H. A., Schneider, E., Larsen, S. & Lo Leggio, L. (2003). *Acta Cryst.* D**59**, 139–141.
Greve, K., La Cour, T., Jensen, M. K., Poulsen F. M. & Skriver, K. (2003). *Biochem. J.* **371**, 97–108.
Hoth, S., Morgante, M., Sanchez, J.-P., Hanafey, M. K., Tingey, S. V. & Chua, N.-H. (2002). *J. Cell Sci.* **115**, 4891–4900.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kikuchi, K., Ueguchi-Tanaka, M., Yoshida, K. T., Nagato, Y., Matsusoka, M. & Hirano, H.-Y. (2000). *Mol. Gen. Genet.* **262**, 1047–1051.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ren, T., Qu, F. & Morris, T. J. (2000). *Plant Cell*, **12**, 1917–1925.
Souer, E., van Houwelingen, A., Kloos, D., Mol, J. & Koes, R. (1996). *Cell*, **85**, 159–170.
Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1991). *J. Cryst. Growth*, **110**, 1–12.
Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1937–1940.
Xie, Q., Frugis, G., Colgan, D. & Chua, N.-H. (2000). *Genes Dev.* **14**, 3024–3036.
Xie, Q., Guo, H.-S., Dallman, G., Fang, S., Weissman, A. M. & Chua, N.-H. (2002). *Nature (London)*, **419**, 167–170.
Xie, Q., Sanz-Burgos, A. P., Guo, H., García, J. A. & Gutiérrez, C. (1999). *Plant Mol. Biol.* **39**, 647–656.